

Journal

für Konflikt- und Gewaltforschung

Journal of Conflict and Violence Research
Vol. 5, 1/2003

Günter Albrecht

Probleme der Prognose von Gewalt durch psychisch
Kranke

[Problems of Predicting Violence by Mentally Ill Persons]

<http://www.uni-bielefeld.de/ikg/jkg/1-2003/albrecht.pdf>

Günter Albrecht

Probleme der Prognose von Gewalt durch psychisch Kranke

1. Vorbemerkung

Jeder, der sich mit Kriminalprognosen beschäftigt hat, weiß, dass sie trotz großer Bemühungen bisher nur bescheidene Leistungen erbringen und die Kriminalpolitiker sowie die Praktiker in der Justiz oft unrealistische Erwartungen in bezug auf die Leistungsfähigkeit von Prognosen haben (vgl. u. a. Andrews 1989). Die Schwierigkeiten und gravierenden Mängel der Gewaltprognose, insbesondere der klinischen Gewaltprognose, sind seit Mitte der 70er Jahre (Monahan 1975; 1981; 1984) nachgewiesen. Daher kann es nur darum gehen, sie durch rationale Handhabung und Einhaltung wissenschaftlicher Standards (vgl. Monahan 1996; Steadman 1987) zu verringern und dem Gesetzgeber klar zu machen, mit welchen „Opfern“ jede Prognosepraxis verbunden ist. Dazu gehören nicht nur die Opfer von nicht-identifizierten „gefährlichen Individuen“, sondern auch die Opfer einer „Prognosepolitik“, die Sicherheit unter allen Umständen maximieren will und durch unangemessene Sicherheitsstandards eine enorme Zahl zu Unrecht Verdächtigter produziert, deren Freiheitsrechte missachtet werden.

Die komplexen Probleme, vor die sich die Prognoseforschung gestellt sieht, vermag der Laie meist nicht zu verstehen. Auch das deutsche Straf- und Strafprozessrecht kündigt von einem fast naiven Vertrauen in die Leistungsfähigkeit von Prognosen, denn es sieht die Erstellung von Gefährlichkeits- und Kriminalprognosen in einer Vielzahl von Rechtsfragen vor (vgl. u. a. Nedopil 2000).

2. Grundprobleme der Kriminalprognose

Die Kriminalprognose (vgl. u. a. Farrington/Tarling 1985a, 1985b; Loeber/Stouthamer-Loeber 1987) dient zur Vorhersage des Auftretens von Kriminalität, entweder in der Form der *Kollektivprognose* oder der Form der Vorhersage der zukünftigen Straffälligkeit einer

bestimmten Person (*Individualprognose*). Auf die letztere müssen wir uns hier beschränken. Bei den *Justizpraktikern* wird zwischen der „Urteilsprognose“ oder der „Entlassungsprognose“ unterschieden, die wiederum abzugrenzen sind von der „Frühprognose“, die schon vor dem Auftreten der Straffälligkeit Aussagen über die zu erwartende Legalbewährung eines Individuums zu machen versucht. Während bei der *Frühprognose* das Handeln des Prognostikers bzw. des auf der Basis der Prognose Intervenierenden *nicht* in die Randbedingungen der Prognose eingeht, beziehen die beiden anderen Prognoseformen ausdrücklich die Überlegung mit ein oder stellen sogar darauf ab, wie sich die *Entscheidung des Prognosestellenden* auf das Verhalten des Individuums auswirken wird. Insofern sind im Strafprozess vorgesehene entscheidungsrelevante Überlegungen immer auch Resultate von wie auch immer zustande gekommenen *Wirkungsprognosen*. Daraus ergeben sich methodologische Probleme, auf die später einzugehen sein wird.

Eine wissenschaftliche Prognose muss sich *erstens* auf eine Theorie zur Erklärung des zu prognostizierenden Verhaltens und *zweitens* auf eine mehr oder weniger vollständige Beschreibung der durch diese Theorie für relevant erachteten Ausgangsbedingungen stützen können (Spieß 1993, 287). Je nach Art der Prognose kennen wir sehr verschiedene Anspruchsniveaus in bezug auf die Präzision dieser „Theorie“ und die Exaktheit der Erfassung der Anfangsbedingungen. Im einzelnen unterscheiden wir zwischen intuitiver, klinischer und statistischer Prognose.

2.1 Intuitive und Klinische Prognose

Die *intuitive* Prognose beruht auf Erfahrungen des Praktikers. Wenig explizite Alltagstheorien treten an die Stelle gesicherter wissenschaftlicher Theorien, der Praktiker operiert mit Faktoren, die auch bei statistischen Prognosen eine Rolle spielen, die er jedoch auf eine ihm selbst nicht klar bewusste, individuelle Weise miteinander verknüpft und gewichtet, ohne sich systematisch über die Richtigkeit seiner Annahmen vergewissern zu können. Volckart (1999, 62) spricht daher davon, dass die intuitive Prognose keine *Methode* ist.

Der intuitiven Prognose nahe stehen die sog. „*klinischen Prognosen*“, die sich darum bemühen, durch „sorgfältig abwägende, verstehende Erfassung und Beurteilung der Täterpersönlichkeit“ (Lefe-

renz 1972, 59) zu prognostischen Aussagen zu kommen, indem sie eine umfangreiche Exploration der Probanden, unter Umständen auch durch Einbezug von testdiagnostischen Untersuchungen und von statistischen Prognosen, vornehmen, wobei die klinische Prognose alle im Einzelfall relevanten Faktoren berücksichtigen und in ihrer Bedeutung für die Kriminalprognose gewichten soll (vgl. Leygraf/Nowara 1992). Statistische und klinische Prognose sind eigentlich keine Gegensätze: „Verschieden ist nur das praktische Vorgehen, weil die einen die in Prognosetafeln geronnenen Erfahrungssätze anwenden und die anderen sich auf ihre ‚subjektive Statistik‘ verlassen“ (Volckart 1999, 62).

Gegen die Präferenz der Justizpraktiker für die klinische Prognose spricht, dass die *klinische* Prognose sich nicht an anderen Populationen bei Kreuzvalidierung *bewähren* kann, da sich der Prognostiker immer nur an seinem *Einzelfall* orientiert und keine expliziten Gründe bzw. quantitativen Operationen benennen kann, die seine Prognose bewirkt haben. Jedes Merkmal wird auf je individuelle Weise in jedem Einzelfall intuitiv gewichtet. Es lässt sich nur *ex post* rekonstruieren, welche Gewichtung vorgenommen wurde, wenn auch ohne um sie zu wissen (vgl. Kühl/Schumann 1989).

Die Forschung verweist auf die gravierenden Fehler klinischer Gewaltprognosen, insbesondere auf den hohen Anteil „falsch Positiver“, also solcher Probanden, denen zu Unrecht Gefährlichkeit attestiert wird. Die Prozentwerte an „*falsch positiv* Prognostizierten“ liegen im *Ausnahmefall* zwischen 54 und 61%, in der Regel jedoch deutlich über 80%, bei *falsch negativen* Werten in derselben Größenordnung (vgl. Monahan 1981; Lab 1988, 101). Man verdächtigt also sehr viele Probanden zu Unrecht, potentieller Gewalttäter zu sein, gleichzeitig gelingt es sehr schlecht, die *tatsächlichen* Gewalttäter zu identifizieren.

2.2 Die statistische Prognose

Die statistische Prognose versucht, in verallgemeinerbarer und überprüfbarer Form jene Merkmale von Probanden zu bestimmen, aus denen auf eine bestimmte Wahrscheinlichkeit künftiger Straftaten geschlossen werden kann (vgl. Clear 1988; Gabor 1986; Tardiff 1989; Tarling/Perry 1985). Bei einer *Konstruktionsstichprobe*, deren Legalbewährung bekannt ist, wird *ex post* ermittelt, welche Eigen-

schaften der Individuen mit Straffälligkeit zusammenhängen: „Der Anteil der unter einer bestimmten Merkmalskombination tatsächlich straffällig gewordenen Personen wird als Maß des individuellen Delinquenzrisikos einer durch diese Merkmale ausgezeichneten Person interpretiert.“ (Spieß 1993, 288) und auf andere Personen generalisiert. Der Prognose unterliegt also ein Syllogismus der folgenden Form: „1. Menschen mit diesen oder jenen Merkmalen handeln unter diesen oder jenen Umständen so und so (nämlich kriminell oder nicht kriminell). 2. Der Proband hat eben diese Merkmale, und er wird voraussichtlich in die gleichen Umstände geraten. 3. Also wird er voraussichtlich ebenso handeln“ (Volckart 1998, 3).

Die Haltbarkeit dieser Hypothese wird anhand einer *Validierungsstichprobe* mit bekannter Legalbewährung geprüft. Es bleibt aber das Problem, ob sich die für die *Vergangenheit* gefundenen Beziehungen auch für die *Zukunft* unterstellen lassen. Durch Erweiterung der einbezogenen Faktoren und induktiv ermittelte Gewichtungen der Merkmale, durch Entwicklung von Strukturprognose-tafeln, die sowohl additive als auch interaktive Effekte von Variablen zulassen (Schultz 1975), kann die Merkmalsverteilung in der *Konstruktionsstichprobe* immer präziser abgebildet und in eine Prognoseformel übersetzt werden. Aber diese *Überanpassung* an die rein zufälligen Besonderheiten der Konstruktionsstichprobe führt bei der Übertragung auf eine *Prognosepopulation* unweigerlich zu einer Schrumpfung der Vorhersagegüte (vgl. u. a. Quinsey 1995).

Die statistische Kriminalprognose beruht auf einer Reihe von pragmatischen Unterstellungen, nämlich zum einen auf den beiden Annahmen, dass biographische Verläufe bestimmten typischen Mustern folgen und dass Personenmerkmale von entscheidender Bedeutung für Verhalten sind, und zum anderen auf den beiden Prämissen, dass die zukünftigen Lebensumstände von Menschen hinreichend gut voraussehbar und dass Rückkoppelungen zwischen Prognose und Verhalten nicht relevant sind. Hinter diesen Annahmen verbergen sich aber virulente Fehlerquellen für Prognosen.

a) *Verlaufsmuster*: Die Identifikation typischer Entwicklungsverläufe bestimmter Kategorien von Menschen ist ein schwer zu lösendes Problem. Die meisten „Karriereuntersuchungen“ basieren auf Daten über *bekannt gewordene* und verurteilte Straftäter, nicht auf *Dunkelfeld*daten. Erstere sind aber eine *verzerrte* Auswahl aller Täter: sie

sind dümmer, weniger gewandt, haben die auffälligeren Taten begangen; sie sind u. U. nur ermittelt worden, weil sie infolge einer psychischen Störung oder als Vorbestrafte bevorzugt Objekt der Aufmerksamkeit der Ermittlungsbehörden werden. Hinzu kommen Selektivitäten nach Alter, Geschlecht, Ethnizität, Schicht etc., so dass es fraglich ist, ob die Kriminologie wirklich weiß, welche Art von Menschen unter welchen Bedingungen kriminelle Handlungen begeht (vgl. Volckart 1998).

Zweitens wird für die Rekonstruktion von Verlaufsmustern aus der Menge der bekannt gewordenen Straftaten, die ja kein verkleinertes Abbild „der“ Kriminalität darstellt, noch eine *verzerrte* Auswahl gezogen, auf die sich die oben erwähnten Erfahrungssätze, insbesondere wenn es um bestimmte Deliktarten bzw. Typen von Straftätern geht, vor allem stützen, nämlich diejenigen, über die ein psychiatrisch-psychologisches Gutachten erstellt wurde. Drittens kann es über Rückfälligkeit keine wirklich *gesicherten* Erfahrungen geben, weil sie in vielen Fällen durch Einsperrern verhindert worden sein könnte, da die bekannt gewordenen Täter ja alle im Rahmen der Strafrechtspflege aufgefallen waren und nur eine Auswahl von ihnen mit minder gewichtigen Taten durch Aussetzung der Vollstreckung in Freiheit geblieben sein dürfte. Bei den Eingesperrten kann niemand sicher sagen, ob sie in dieser Zeit rückfällig geworden *wären*, wenn auch sie in Freiheit hätten verbleiben können.

Viertens ist davon auszugehen, dass auch kriminalprognostische Erfahrungssätze *zeit-* und *gesellschaftsabhängig* sind. Ihre Validität dürfte daher mit zeitlicher und gesellschaftlicher Entfernung zwischen der Generierung dieses Erfahrungswissens und seiner Anwendung abnehmen (vgl. Volckart 1998, 4). Andererseits sind kurz- und mittelfristige Prognosen wenig brauchbar. Um zu (relativ!) sicheren Datengrundlagen zu kommen, benötigt man in der empirischen Forschung möglichst viele Fälle, die insbesondere bei den gravierenden Delikten, die erfreulicherweise relativ selten sind, aber erst nach längerer Zeit zusammenkommen. Muss man *lebenslange prospektive Verlaufsforschungen* anstellen, so liegen die entsprechenden Ergebnisse erst nach mehreren Jahrzehnten vor. Dann aber stellt sich die Frage, ob *diese* Ergebnisse noch auf die Gegenwart, bzw. genauer: die Zukunft, und die neue Population übertragen werden können, denn die Rahmenbedingungen haben sich wäh-

rend dieses langen Zeitraumes unter Umständen sehr gravierend verändert.

b) *Die Merkmale des Probanden:* Die oben angesprochenen Gesetzmäßigkeiten des Handelns stellen immer auf die Zugehörigkeit der fraglichen Individuen zu einer bestimmten *Kategorie von Personen* mit ganz bestimmten *Persönlichkeitseigenschaften* ab. Die Abklärung dieser beschränkten Zahl von Merkmalen aus der unendlichen Fülle menschlicher Eigenschaften erfolgt im Kontext des Strafprozesses bzw. der dabei anfallenden Ermittlungen. Geht es um die Erfassung von Persönlichkeitseigenschaften oder Diagnosen, so ist diese an die Kooperation des Probanden gebunden. Aus der empirischen Sozialforschung ist gesichert, dass Befragte eine starke Tendenz aufweisen, ihr Antwortverhalten an dem zu orientieren, was sie für sozial erwünscht halten, auch wenn sie im Prinzip nicht viel zu verlieren haben (vgl. Albrecht 1999, 785ff; Mummendey 1981; Phillips 1973). Personen aber, die wissen, dass vom Verlauf eines Strafprozesses ihr weiteres Leben in entscheidender Weise beeinflusst werden dürfte, haben ein sehr großes Interesse, durch eine kalkulierte Selbstdarstellung bei relevanten Dritten einen Eindruck von sich selbst zu erzeugen, der den Prozessverlauf möglichst günstig zu gestalten verspricht. Die Informationen, die Probanden von sich im Rahmen eines Prozesses zu verbreiten versuchen, sind daher von zweifelhafter Validität.

c) *Die erwarteten Lebensumstände:* Prognosen stellen nicht nur darauf ab, dass bestimmte, mehr oder weniger fixe Merkmale der Person mit einer gewissen Notwendigkeit eine bestimmte Wahrscheinlichkeit des Auftretens/Wiederauftretens eines problematischen Verhaltens bedingen, sondern sie binden diese Annahme an bestimmte zukünftige Lebensbedingungen dieser Person. „Diese Prognosen der Lebensumstände, in denen der Proband im Falle seiner Freilassung voraussichtlich leben würde, sind prinzipiell unsicher, und sie sind es umso mehr, je weiter sie in der Zukunft liegen“ (Volckart 1998, 5), es sei denn sie ergäben sich eindeutig aus der Prognose selbst, z. B. dadurch, dass der Prognostiker den Probanden wegen seiner vermeintlichen Gefährlichkeit hinter Schloss und Riegel hält.

d) *Rückkoppelung:* Damit wären wir bei der unter Umständen ganz besonders problematischen, vierten Art von Fehlerquellen, den Rückkoppelungseffekten zwischen Prognose und Verhalten. Zu unterscheiden sind für unsere Problematik 4 Varianten von Rückkoppelungseffekten. 1. Eine zunächst *falsche ungünstige* Prognose wird dadurch, dass sie gestellt wird, zur *richtigen ungünstigen* Prognose (ein Proband, der an sich nicht straffällig geworden wäre, sieht sich durch die negative Prognose so weit degradiert, dass er eine vorher nicht vorhandene Motivation zu kriminellen Handeln entwickelt). 2. Eine zunächst *falsche günstige* Prognose kann dadurch, dass sie gestellt wird, zur richtigen werden: Das gewährte Vertrauen führt zu einer moralischen Selbstverpflichtung und/oder neuem Selbstkonzept, so dass die an sich gegebene Disposition zu kriminellen Handeln beseitigt wird. 3. Eine zunächst *richtige ungünstige* Prognose kann dadurch, dass sie gestellt wird, u. U. zur *falschen ungünstigen* Prognose werden, z. B. dann, wenn der Proband den Prognostiker Lügen strafen will. 4. Eine zunächst *richtige günstige* Prognose kann dadurch, dass sie gestellt wird, zur *falschen günstigen* Prognose werden; sei es, dass bei subjektivem Strafbedürfnis des Probanden eine Strafaussetzung zur Verachtung der Justiz und der Gesellschaft führt oder das ungerechtfertigte Vertrauen in die Fähigkeit des Probanden, in Zukunft „sauber zu bleiben“, diesen zu leichtfertigen Verhaltensweisen verleitet, die dann kriminelles Verhalten verursachen.

Rückkoppelungen dieser Art können *nie* ausgeschlossen werden und man kann nicht wissen, welcher Effekt bei wem eintreten wird.

3. Probleme der Bewertung der empirischen Bewährung statistischer Prognosen

Wenn Prognosen schon nicht perfekt sein *können*, dann muss man umso mehr Wert darauf legen, dass man unter verschiedenen zur Verfügung stehenden Prognoseinstrumenten jene auswählt, die sich noch am besten bewähren. Es stellt sich aber als sehr schwer heraus, angemessene *Vergleichsmaßstäbe* zu finden, denn Prognosen können aus mathematischen Gründen nicht unter allen Bedingungen gute Leistungen erbringen. Eine der entscheidenden Voraussetzungen für gute Prädiktionsleistungen ist die Höhe der „Basisrate“

des zu prognostizierenden Verhaltens. Das soll in Anlehnung an ein Beispiel von Kühl/Schumann (1989, 131f) illustriert werden.

Nehmen wir das Beispiel der Tötungsdelikte, die ja relativ selten begangen werden, und unterstellen einmal eine *Basisrate* von 0,1% – was heißen würde, dass unter 1000 Personen eine eine andere töten würde. Wollte man die potentiellen Täter durch ein geeignetes Prognoseinstrument identifizieren, um sie vorbeugend zu verwahren und/oder zu behandeln, und hätte zu diesem Zweck einen Test von ganz *außergewöhnlicher* Treffergenauigkeit, nämlich 95%, entwickelt und würde diesen Test bei 100 000 Probanden anwenden, dann würden von den 100 Personen, die angesichts der oben benannten Basisrate als potentielle Täter zu erwarten wären (eben diese 0,1%), 95 durch den Test als potentielle Täter identifiziert. Obwohl die Fehlerrate mit 5% sehr niedrig liegt, würden aber von den 99 900 übrigen Personen, die keine wirklichen potentiellen Täter sein sollten, 5% *fälschlicherweise* als potentielle Täter ausgewiesen, also ca. 4 995 Personen. Würde man auf der Basis dieses Testes eine vorbeugende Haft vornehmen, müsste die Einsperrung von 95 tatsächlichen potentiellen Tätern mit dem sehr bedenklichen Nebeneffekt erkaufte werden, dass zugleich 4 995 Personen, von denen in Wirklichkeit keine Gefahr ausgeht, zu Unrecht eingesperrt würden. Sähe man die vorbeugende Haft für tatsächlich Gefährliche für legitim an, so läge das Verhältnis von illegitim zu legitim vorbeugend Inhaftierten bei ca. 53 zu 1. Die Frage stellt sich, ob man ein solches Missverhältnis ethisch vertreten kann. Selbst wenn man die Treffergenauigkeit auf unrealistische 99% hinaufschrauben könnte, ergäben sich bedenkliche Relationen. Bei 100 000 Untersuchten und einer Basisrate von 0,1% wären es 100 Gefährliche, von denen nun 99 identifiziert würden. Andererseits würden von den 99 900 Ungefährlichen auch jetzt 1% falsch prognostiziert, also als gefährlich eingestuft, obwohl sie es nicht sind. Das wären immerhin noch 999 Personen, und auf einen legitim eingesperrten Menschen kämen immer noch 10 illegitim weggesperrte Menschen.

Das Problem liegt also bei der niedrigen *Basisrate*, die die Effekte der guten Vorhersagegenauigkeit unterläuft und durch diese kaum entscheidend ausgeglichen werden kann. Dabei ist zu beachten, dass die Treffergenauigkeit bei beiden Beispielen als ganz unrealis-

tisch hoch angesetzt worden ist, wie an anderer Stelle zu zeigen sein wird. Die *Treffergenauigkeit* (= Summe der wahr Positiven und der wahr Negativen in Relation zur Testhäufigkeit) ist in der Regel eine Funktion der Wahl des Testtrennwertes, des sog. „Cutting Points“ (vgl. Amelang/Zielinski 1997, 364), den der Prognostiker wählt. Dies ist der Punkt auf einer zur Prognose verwendeten Skala, ab dem man einen Probanden als zur Risikogruppe gehörend klassifiziert. Durch Variation dieses Cutting Points kann man die Treffergenauigkeit variieren, erkaufte das aber mit ganz spezifischen Risiken, z. B. einem prekären Verhältnis von falsch Positiven zu wahr Positiven etc.: Je niedriger man den Cut Off wählt, desto sicherer kann man sein, die wahr Positiven zu entdecken, aber man muss auch mit erhöhten Anteilen von falsch Positiven rechnen. Will man extrem sicher sein, alle Positiven zu identifizieren, kann man das nur mit einem außerordentlich hohen Risiko von falsch positiven Klassifikationen erkaufen.

Erschwerend kommt hinzu, dass die angesprochenen Kennziffern in den meisten Fällen den Instanzen, die Prognosen erstellen sollen, gar nicht wirklich zur Verfügung stehen. Dies gilt insbesondere für die Basisrate, denn im Normalfall werden ja die problematischen Probanden nicht entlassen, können also nicht rückfällig werden, aber auch nicht zeigen, dass sie auch in Freiheit *nicht* rückfällig geworden wären.

Als prekär gelten für die Öffentlichkeit in der Regel die *falsch Negativen*, also die Personen, die *tatsächlich* gefährlich sind, aber durch den Test *nicht* entdeckt werden, eben fälschlicherweise einen „negativen Befund“ im Sinne der medizinischen Terminologie aufweisen.¹ Wie hoch die Rate falsch Negativer ist, schwankt selbstverständlich erheblich; sie wird oft in der Öffentlichkeit erheblich überschätzt. Sie liegt z. B. beim Problem der Strafaussetzungen bei Lebenslänglichen im ersten Jahr unter 1% (vgl. Weber 1988, 106). Die repressive Funktion der Toleranz einer hohen Rate falsch Positiver ist schwer zu belegen, da die restriktive Prognostik nur im Ausnahmefall dokumentierbar ist: der falsch positiv klassifizierte Gefangene

¹ In Anlehnung an den anglo-amerikanischen Sprachgebrauch werden im Text folgende Abkürzungen verwendet: FN = falsch Negative (Individuen, die fälschlich als „ungefährlich“ prognostiziert wurden), FP = falsch Positive (Individuen, die fälschlich als „gefährlich“ prognostiziert wurden), TN = wahre Negative (Individuen, die richtig als „ungefährlich“ prognostiziert wurden), TP = wahre Positive (Individuen, die richtig als „gefährlich“ prognostiziert wurden).

bleibt in Haft oder der vermeintlich gefährliche Patient bleibt in der Psychiatrie – und wenn man ihn nach Jahrzehnten doch freilässt und er im hohen Alter ohne Delikte bleibt, wird dies – Ironie der Geschichte – noch als Erfolg der Haft oder Behandlung gedeutet.

Dass die Zahl der falsch Positiven erschreckend hoch sein kann, ließ sich im Verlauf eines unfreiwilligen Realexperiments in den forensischen Kliniken der USA zeigen, als diese infolge einer durch das oberste Gericht angeordneten Haftüberprüfung aller psychisch kranken Rechtsbrecher Tausende entlassen mussten und sich in der Folge ergab, dass für 97,4% der als „gefährliche psychisch kranke Straftäter“ eingesperrten Probanden diese Unterbringung als unangemessen angenommen werden muss, da sie nach ihrer erzwungenen Freilassung nicht einschlägig rückfällig wurden – was ihnen zuvor ja auf Dauer unterstellt worden war (vgl. zu diesem berühmten Baxtrom-Fall u. a. Cocozza 1975; Cocozza/Steadman 1976; Steadman/Cocozza 1974; 1978).

Aus den genannten Gründen gilt ferner, dass die *Auswahlrate* (Summe von wahr Positiven und falsch Positiven im Verhältnis zur Testhäufigkeit) entscheidend für eine mögliche Treffsicherheit der Prognose ist. Höchste Treffsicherheit kann sich nur ergeben, wenn Auswahlrate und Basisrate identisch sind. Ist die Zahl falscher Negativer und falscher Positiver jeweils = 0 (was ja ideal wäre), dann ergibt sich die Basisrate aus der Relation der Zahl der wahren Positiven zur Summe der wahren Positiven und der wahren Negativen. Jede Veränderung der Auswahlrate muss in diesem Fall die Treffsicherheit vermindern. Da die Auswahlrate oft politisch vorgegeben wird und nicht an der meist gar nicht bekannten Basisrate orientiert ist, muss man damit rechnen, dass die Fehlerquote zu Lasten der falschen Positiven verschoben wird.

Die Prognoseleistungen von verschiedenen Verfahren zu vergleichen, erweist sich also als erheblich schwieriger, als der Laie ahnt (vgl. Mossman 1994). Eine Kennziffer, die eine angemessene Bewertung erlaubt, ist die Verbesserung gegenüber der Zufallszuweisung (RIOG). Hier ist die sog. ROC-Kurve (Receiver Operating Characteristics) aus der Signalerkennungstheorie hilfreich, die die Beziehung zwischen der Erkennung eines Signals und den sog. „falschen Alarmen“ bei verschiedenen Entscheidungsschwellenwerten beschreibt. Entsprechend kann man die sog. „Fläche unter der Kurve“ („Area Under Curve“; AUC) berechnen, die sich ergibt, wenn

man die Werte der Spezifität² (bzw. 1 minus Spezifität) und der Sensitivität³ miteinander verknüpft, die sich für bestimmte Schwellenwerte (Cut offs) einer Prognoseskala ergeben: Der erste Wert ist niedrig, wenn die Spezifität sehr groß ist. Wenn schon bei niedrigen Werten der ersteren (also dann, wenn die erwünschte Spezifität hoch ist) auch hohe Werte für die Sensitivität vorliegen, was natürlich erwünscht ist, dann wird die Fläche unter der Kurve, die sich für verschiedene Schwellenwerte (cut offs, die ja mit der Selektionsrate identisch sind) bestimmen lässt, größer, als wenn eine hohe Spezifität mit niedriger Sensitivität einhergeht. Die „Fläche unter der Kurve“ ist ein Qualitätsmerkmal der Prognoseleistung, das *nicht abhängig von der Basisrate* ist, wie z. B. die falsch und wahr positiven Raten etc. (Mossman 1994; Rice, 1997; Rice/Harris 1995; vgl. Faust/Nurcombe 1989; vgl. Amelang/Zielinski 1997, 368). Dieser AUC-Wert schwankt zwischen 0 (perfekt negative Vorhersage) über .50 (Zufallszuordnung) bis 1.0 (perfekte positive Vorhersage). Die Werte können wie folgt interpretiert werden: Der AUC-Wert entspricht der Wahrscheinlichkeit, dass eine zufällig ausgewählte tatsächlich gewalttätige Person höhere Punktwerte auf dem Instrument aufweist als eine zufällig ausgewählte tatsächlich nicht gewalttätige Person (Mossman 1994; Rice/Harris 1995). Nennenswerte AUC-Werte kann man ab ca. .70 annehmen.

4. Neuere Prognosestudien zur Gewalt durch psychisch kranke Rechtsbrecher

Die Prognose der Gewaltkriminalität von psychisch Kranken stellt ein besonderes Problem dar, weil in der Öffentlichkeit eine starke Tendenz vorherrscht, psychisch Kranke generell für besonders gefährlich zu halten; ein Vorurteil, das – so Nedopil (1999) – sehr zu Lasten der psychisch Kranken geht. Es resultiert nicht zuletzt daraus, dass in der Mehrzahl von Kriminalfilmen die besonders schweren Gewaltdelikte von psychisch kranken Personen begangen werden, bei denen die scheinbar enorme Zwanghaftigkeit des gewalttätigen Handelns beim Rezipienten die Vorstellung erzeugt, von der-

² Die Spezifität ist wie folgt definiert: $Sp = TN/(FP + TN)$

³ Die Sensitivität ist wie folgt definiert: $Sens = TP/(TP + FN)$

artigen Tätern gehe eine permanente und nie endende Gefahr für andere Menschen aus.

Die Forschung belegt, dass es sich tatsächlich um ein Vorurteil handelt. So zeigten z. B. Steadman u. a. (1998), dass es keine signifikante Differenz in der Gewaltprävalenz zwischen entlassenen psychisch Kranken und der „Normalbevölkerung“ in den Wohngebieten gibt, in die die Patienten entlassen wurden, wenn man die Variable „Substanzmissbrauch“ kontrolliert. Die Mehrheit der Gewalttätigkeiten beider Teilgruppen richtete sich zudem gegen Familienmitglieder und Freunde und fand zu Hause statt. Die gezeigte Gewalt trifft in der Regel also nahestehende Personen, nicht „die Allgemeinheit“ (so auch Nedopil 1999, 123). Es besteht daher eigentlich kein Grund, der Gewalt durch psychisch Kranke *insgesamt* besondere Aufmerksamkeit zu widmen. Dabei ist natürlich zu beachten, dass die Prognose der Gewalt von psychisch kranken Gewalttätern ein *spezielles*, aber besonders brisantes Problem darstellt. Die Gesetzgeber und die Öffentlichkeit sehen offensichtlich nur diese Teilgruppe. Daher seien im folgenden einige neuere Studien zur Gewaltprognose bei dieser Population vorgestellt.

Die Prognose der Gewalt und der „Gefährlichkeit“ von psychisch kranken Rechtsbrechern stellt ein besonderes Problem dar (vgl. u. a. Black/Spinks 1985), das in der Vergangenheit – häufig verfassungswidrig – zu Lasten der psychisch Kranken durch eine ungerechtfertigt restriktive Beschränkung der Freiheit dieses Personenkreises gelöst worden ist (s. o.). Die Schwierigkeiten liegen zum einen darin, dass wegen der Affinität psychisch Kranker zum klinischen Milieu der Rückgriff auf *klinische* Prognosen nahe liegt, obwohl sie statistischen Prognosen unterlegen sind (vgl. u. a. Gottfredson 1987). Zum anderen sind die meisten Prognoseinstrumente für *Kriminalität insgesamt* entwickelt worden, deren Prognose (Vorhersagegenauigkeit für Kriminalität zwischen 60 und 80%; vgl. Andrews 1989; Holland u. a. 1983; Nuffield 1982) in der Regel besser gelingt als die von Gewalttaten. Dies dürfte an den niedrigen Basisraten liegen (Rückfall bei Gewalttätern zwischen 16 und 77%; vgl. Quinsey 1980; Steadman 1987), die bedingen, dass es kaum möglich ist, eine bessere Prognose als die zu stellen, *keiner* werde rückfällig. Das Verhalten von psychisch Kranken ist noch schwerer zu prognostizieren – unter Umständen weil diese noch sensibler auf die soziale Umwelt und ihre Turbulenzen reagieren.

Von besonderem Interesse ist für die Öffentlichkeit natürlich die Prognose des Verhaltens psychisch kranker Rechtsbrecher *außerhalb* der psychiatrischen Anstalt bzw. der forensischen Klinik.

4.1 Gewaltrückfall außerhalb der Anstalt

a) *Die Entwicklung des Violence Risk Appraisal Guide (VRAG)*: Harris u. a. (1993) versuchten, ein statistisches Prognoseinstrument zu entwickeln, das sowohl für Rückfalltäter generell als auch für psychisch kranke Rechtsbrecher anwendbar ist. Daher wurden zwei entsprechende Teilgruppen einbezogen, nämlich zum einen 371 Männer, die in eine Hochsicherheitspsychiatrieeinrichtung eingewiesen worden waren, und zum anderen 324 verurteilte Männer, die ein kurzes psychiatrisches Screening durchlaufen hatten. Die Individuen der beiden Gruppen wurden miteinander in Bezug auf zentrale Kriterien (Deliktart, Deliktsschwere, frühere Häufigkeit und Schwere gewalttätigen Verhaltens und nicht-gewalttätigen deliktischen Handelns, Alter, Zeitpunkt des zur jetzigen Verurteilung führenden Deliktes) „gematched“. Probanden, die aufgrund der andauernden Einsperrung keine Gelegenheit gehabt hätten, rückfällig zu werden, wurden ausgeschlossen, so dass noch 618 Probanden übrig blieben.

Die Probanden waren fast alle *schwere Gewalttäter*, die restlichen hatten ebenfalls meist schon mehrfach Gewaltdelikte verübt. Viele Probanden hatten Tötungsdelikte begangen und relativ viele bei bedingter Entlassung versagt. Viele wiesen psychopathologische Störungen auf und hatten schwere Alkoholprobleme. Beide Teilpopulationen verbrachten anschließend mehrere Jahre entweder in Behandlung in der geschlossenen Psychiatrie oder im Strafvollzug. Der mittlere zeitliche Abstand zwischen dem „Index-Delikt“, dessenwegen sie verurteilt worden waren, und der ersten Gelegenheit zum Rückfall betrug 5,31 Jahre, wobei für die Rückfallwahrscheinlichkeit kaum bedeutsam war, ob die Probanden in Behandlung waren oder dem Strafvollzug unterlagen.

Alle Variablen, außer der, die sich auf den Rückfall bezog, wurden dem gesamten klinischen Material und den Strafprozessakten ex post entnommen (u. a. Fehlanpassung im Kindergarten, Weibliche Opfer der begangenen Gewalttaten, Antisoziale Persönlichkeitsstörung nach dem DSM-III auf der Basis von Problemverhalten in

der Kindheit,⁴ Schizophrenie nach dem DSM-III, Ehelosigkeit, Alkoholabusus in der Jugend, Trennung von den Eltern vor dem sechzehnten Lebensjahr, Aggression im Erwachsenenalter, Ausmaß der Überwachung, Karriere als Eigentumskrimineller, Grad der Verletzung des Opfers, Psychopathie nach dem PCL-R (Hare 1991)).

Die zur Messung der „Psychopathie“ verwendete *Revised Psychopathy Checklist* (PCL-R) (20 Items) soll trotz ihrer ursprünglich anderen Zielsetzung auch allein brauchbare *prädiktive* Validität besitzen (vgl. Harris u. a. 1991). Sie umfasst eine Subskala zur Erfassung emotionaler und interpersonaler Defizite (Gefühllosigkeit, emotionale Verflachung, mangelnde Schuldeinsicht, Fehlen von Empathie, Manipulationstendenzen) und eine zur Messung problematischer Verhaltensweisen (schwache Verhaltenskontrolle, parasitärer Lebensstil). Auch die Items dieser Skala wurden auf der Basis der vorhandenen Akten- und Probandenunterlagen erfasst.

Ein unabhängiges Team erfasste die Rückfälligkeit auf der Basis der Daten aller in Frage kommenden Kontrollinstanzen. Rückfälle wurden als *gewalttätig* klassifiziert, wenn sie eine neue Anklage wegen eines Gewaltverbrechens auslösten oder zu einer erneuten Unterbringung in einer Hochsicherheitseinrichtung wegen eines Gewaltverbrechens führten. Als *Gewaltdelikte* galten alle Körperverletzungen und Sexualverbrechen mit Gewalt, bewaffneter Raub, gewalttätige Freiheitsberaubung, Bedrohung/Nötigung und Einsetzen/Vorzeigen einer Feuerwaffe (nicht ihr Besitz), Raub und Brandstiftung. Eine Person wurde als über eine Rückfallgelegenheit verfügend definiert, wenn sie in *Freiheit* entlassen oder in den *offenen* Strafvollzug oder in eine *offene* psychiatrische Station verlegt wurde.

Da sich die verwendeten Prognosevariablen für die als mental krank definierte Population in gleicher Weise bewährten wie bei den nicht als psychisch krank Klassifizierten, wurden die Gruppen zusammengefasst. Für die Auswahl der Prädiktoren wurden schrittweise Diskriminanzanalysen für jeden der vier Variablenblöcke (Kindheitsgeschichte, Anpassung im Erwachsenenalter, Index-Delikt, Assessmentwerte) und jede der fünf Teilpopulationen (a) Gesamtsample, (b) behandelte Probanden, (c) getestete Probanden,

⁴ Gemeint ist das „Diagnostic and Statistical Manual of Mental Disorders“ der American Psychiatric Association in seiner dritten Auflage; vgl. Koehler/Saß 1984.

(d) Zufallsauswahl im Umfang der halben Gesamtgruppe und (e) Rest des Samples berechnet.

Die durchschnittliche Risikodauer⁵ für alle Probanden lag bei ca. 81 Monaten. Insgesamt waren 31% der Probanden mit Gewaltdelikten i. o. S. rückfällig. Die univariaten Korrelationen zwischen Prädiktoren und Rückfall waren noch am höchsten bei der Psychopathy Checklist ($r = .34$), bei Fehlanpassung im Grundschulalter ($r = .31$), beim Alter beim Eintritt in die kriminelle Karriere ($r = -.26$), DSM-III Personality Disorder ($r = .26$), Trennung von den Eltern unter 16 Jahren ($r = .25$) und Scheitern bei früheren Entlassungen ($r = .24$) – alles aber sehr bescheidene Korrelationen.

Da die sich als relevant erweisenden zwölf Variablen in der Diskriminanzanalyse je nach Untersuchungspopulation schwankten, wurden jene Variablen eliminiert, die nur in zwei oder weniger Modellen eine signifikante Zunahme der Erklärungsleistung erbrachten. Für die Berechnung der Skalen wurden drei Varianten der Handhabung der Variablen praktiziert, nämlich zum einen eine ungewichtete Verrechnung, zum zweiten eine Gewichtung entsprechend ihrem Beitrag zur Diskriminanzfunktion und drittens die Verwendung der Nuffield-Methode (Nuffield 1982), bei der jede Variable anhand des Kriteriums, wie viel Prozent der Differenz zum Mittelwert des Rückfalls eine spezifische Variablenausprägung bedingt, gewichtet wird.

Für die prädiktive Validität des endgültigen Instruments ergeben sich für 5 Abstufungen (Cut off) des Risikoscores (20. Perzentil, 40. Perzentil, 60. Perzentil, Basisrate und 80. Perzentil) die in Tabelle 1 aufgeführten Kennziffern. Es zeigt sich der typische Befund, dass sich dann, wenn die *Auswahlrate* (festgelegt durch den Cut off-Wert) in etwa der *Basisrate* entspricht, ein passabler Ausgleich zwischen den Kriterien „Prozentanteil der korrekt Klassifizierten“ (73%), der „Rate der falschen Positiven“ (45%), der „Sensitivität“ ($TP/(TP+FN)$) (60%) und der „Spezifität“ ($TN/(FP+TN)$) (78%) einstellt. Es wird aber deutlich, wie begrenzt die Leistungsfähigkeit der Prognose ausfällt: Über einen Anteil der korrekt Klassifizierten von 74% kommt man auch hier nicht hinaus. Selbst bei einem Cut off, der der Basisrate entspricht, müssen wir bei diesem Instrument mit einer Rate der falschen Positiven von 45% rechnen – und das In-

⁵ Dauer der Zeit, in der ein Proband wegen der geringeren Kontrollen Gelegenheit gehabt hätte, Gewaltdelikte zu begehen; siehe Ende des vorletzten Absatzes.

strument identifiziert dennoch nur 60% derjenigen, die rückfällig werden.

Tabelle 1: Prognoseleistungen nach Harris u. a. (1993)

	20. Perzentil	40. Perzentil	60. Perzentil	Basisrate	80. Perzentil
Prozentanteil der korrekt Klassifizierten	47%	61%	70%	73%	74%
Falsch Positive	64%	57%	49%	45%	39%
Sensitivität	95%	87%	72%	60%	41%
Spezifität	25%	49%	69%	78%	88%

Ein Rückfall kann natürlich nur bei denen eintreten, die eine *Gelegenheit* dazu haben, die also *entlassen* worden sind. Ein Vergleich zwischen Entlassenen und den Nicht-Entlassenen zeigte, dass die Nicht-Entlassenen in der Vorgeschichte die schwereren Delikte begangen hatten, obwohl andererseits die Schwere des Deliktes negativ mit der Rückfallwahrscheinlichkeit korrelierte.

Unter dem Gesichtspunkt der *Generalisierbarkeit* der Befunde ist zu berücksichtigen, dass viele der Probanden zwar als psychisch kranke gefährliche Gewalttäter galten und sicher auch einige von ihnen psychotisch waren, sich aber dennoch die Frage stellt, ob sie sich grundsätzlich von Gewalttätern in Gefängnissen unterscheiden. Zu beachten ist, dass ein großer Anteil *nicht* als psychisch krank galt (55%) und dass von den als psychisch krank geltenden Probanden die Mehrheit (58%) *nicht* als *psychotisch* angesehen wurde, sondern nach Meinung der Gutachter als durch schwere „antisoziale Störungen“ belastet galt. Zu bedenken ist auch, dass zwar nach dem DSM-III ca. 30% der Probanden Psychotiker waren, dass aber andererseits unter den üblichen Gefängnisinsassen ebenfalls ca. 25% als Psychotiker eingestuft werden (Hodgins 1990) und dieser Prozentanteil unter Insassen mit sehr schweren Delikten noch höher sein dürfte; was für die Generalisierbarkeit spricht. Es empfiehlt sich, Studien zu *spezifischen Tätergruppen* durchzuführen, da sich die Rückfallverläufe für besondere Tätergruppen sehr unterscheiden (vgl. z. B. die Unterschiede zwischen Kinderschändern und Gewalttätern mit erwachsenen Frauen als Opfern).

Die Prognoseleistungen dieses mit zwölf Items sehr kurzen Prognoseinstruments von Harris u. a. – später VRAG (Violence Risk Assessment Guide) genannt – erwiesen sich bei verschiedenen Populationen in bezug auf gewalttätigen Rückfall als nicht schlecht. So berichteten Rice/Harris (1995) AUC-Werte zwischen .73 und .75, während sie bei der Prognose von Sexualstraftätern weniger gute Ergebnisse erreichten (Rice/Harris 1997). Hier sei daran erinnert, dass der AUC-Wert die Wahrscheinlichkeit angibt, dass eine zufällig ausgewählte tatsächlich gewalttätige Person höhere Punktwerte auf dem Prognoseinstrument aufweist als eine zufällig ausgewählte tatsächlich nicht gewalttätige Person. AUC-Werte von .73 bis .75 bedeuten also gegenüber der Zufallszuweisung (.50) eine Verbesserung von 23 bis 25%. Da zunächst die *dynamischen* Aspekte des Akteurs ausgeblendet blieben, haben die Autoren später klinische Variablen einbezogen und den VRAG zum *Violence Prediction Scheme* weiterentwickelt (Webster u. a. 1994; Quinsey u. a. 1998).

b) Historical, Clinical, and Risk Management Instrument (HCR-20), VRAG und PCL-R im Vergleich: Ein anderes neues Prognoseinstrument stellt das sogenannte *Historical, Clinical, and Risk Management (HCR-20)-Gewalttrisiko-Assessment Instrument* dar, das u. a. *rückblickend* von Douglas/Webster (1999) auf ein Sample von 75 männlichen kanadischen, von Bundesgerichten zu Hochsicherheitshaft verurteilten psychisch kranken Rechtsbrechern angewendet wurde, um seine Eignung mit anderen Instrumenten zu vergleichen. Dabei ist zu beachten, dass es nicht um eine Prognose im *eigentlichen* Sinne geht, sondern darum, inwieweit man auf der Basis der Informationen über die Vergangenheit bzw. die Karriere der Probanden, die in diese Instrumente eingehen, die Delikte hätte prognostizieren können, von denen man zum gegenwärtigen Zeitpunkt weiß, dass sie diese bisher begangen haben.

Diese aktuarielle⁶ Skala setzt sich aus drei Teilskalen zusammen. Die Teilskala *Historical Scale* umfasst zehn *statische* Items (frühere Gewalt, erstes Gewaltdelikt im frühen Alter, Instabilität in sozialen Beziehungen, Beschäftigungsprobleme, Substanzmissbrauch, ernste mentale Erkrankung, Psychopathie, frühe Fehlanpassung, Persön-

⁶ Von aktuariellen Instrumenten spricht man im anglo-amerikanischen Sprachraum, wenn es sich um eine Prognosetafel handelt, bei der auf der Basis vorhandener Daten eine streng quantitative Risikoabschätzung vorgenommen wird.

lichkeitsstörung, früheres Bewährungsversagen). Die sogenannte *Clinical Scale*-Teilskala besteht aus fünf dynamischen und variablen Items (mangelnde Einsicht, negative Einstellungen, aktive Symptome von ernsten psychischen Störungen, Impulsivität, Nichtansprechen auf Behandlung), während die Teilskala „Risk Management Scale“ fünf Items zu den zukünftigen Lebensplänen des Probanden umfasst (Mangelnde Realisierbarkeit, Exposition an destabilisierende Einflüsse, mangelnder persönlicher Support, Noncompliance in Bezug auf Wiedereingliederungsversuche, Stress). Des weiteren kamen die *PCL-R* (Hare 1991) (s. o.) und der *VRAG* (s. o.) (Harris u. a. 1993; Harris/Rice 1997; Rice 1997; Rice/Harris 1997) zur Anwendung.

Bei der Analyse der univariaten Korrelationen zwischen dem HCR-20, dem VRAG und der PCL-R auf der einen und früheren Verurteilungen auf der anderen Seite wird für die verschiedenen Delikte für die einzelnen Skalen bzw. Subskalen ersichtlich, dass die sich auf die Daten der Vergangenheit stützende Teilskala HCR-20 Historical für alle Deliktarten die deutlich höheren Korrelationen aufweist (vgl. Tab. 2). Es zeigt sich ferner, dass *keine* der Skalen signifikant mit *Sexualdelikten* verknüpft ist, wohl aber mit *reinen* Gewaltdelikten und Delikten *insgesamt*.

Tabelle 2: Prognoseleistungen (Korrelationskoeffizienten) des HCR-20, VRAG und des PCL-R nach Douglas/Webster (1999) (in Klammern: Irrtumswahrscheinlichkeit)

	Alle Delikte	Gewaltdelikte, inkl. Sexualdelikte	Gewaltdelikte (nur nicht-sex.)	Sexualdelikte
HCR-20 Historical	.47 (.001)	.50 (.001)	.53 (.001)	.06 (ns)
HCR-20 Clinical	.16 (ns)	.30 (.01)	.25 (.05)	.16 (ns)
HCR-20 Total	.33 (.01)	.42 (.001)	.44 (.001)	.13 (ns)
VRAG	.44 (.001)	.28 (.01)	.20 (ns)	-.08 (ns)
PCL-R	.35 (.01)	.41 (.001)	.38 (.001)	.15 (ns)

Bei einer schrittweisen multiplen Regression ergibt nur der „historische“ Part von HCR-20 eine signifikante „Prädiktion“. Bei einer hierarchischen Regression mit erzwungenem Einschluss klärt die HCR-20-Skala zusätzliche Varianz auf, wenn die Werte der PCL-R-Skala zuvor schon eingeschlossen waren, aber nicht umgekehrt.

Ähnliches ergibt sich beim Vergleich mit der VRAG-Skala: Wenn die Werte der HCR-20-Skala schon eingeschlossen waren, dann erklären die Werte des VRAG darüber hinaus nichts, während umgekehrt die später eingeschlossene HCR-20-Skala noch einen deutlich größeren Anteil der Varianz zusätzlich erklärt als zuvor die VRAG-Skala allein.

Nicht-parametrische Vergleiche zwischen den verschiedenen dichotomen Werten gewalttätigen oder anderweitigen „antisozialen Verhaltens“ für Probanden mit Werten unterhalb des Medians und denen mit Werten oberhalb des Medians des HCR-20 bzw. der jeweiligen Teilskalen und eine Berechnung der odds ratios erbrachten interessante Befunde: Die odds ratios geben an, mit welcher Wahrscheinlichkeit ein Proband, der in früherer Zeit wegen eines Gewaltdelictes verurteilt wurde, verglichen mit einer Person ohne diese Vorgeschichte einen Wert oberhalb des Medians des HCR-20 aufweist. Umgekehrt kann man sagen, dass dieser Wert die Wahrscheinlichkeit angibt, mit der eine Person mit Werten oberhalb des Medians des HCR-20 eine entsprechende deliktische Vorbelastung aufweist. Während in der Literatur odds ratios von ca. 2,5 als substantiell und klinisch bedeutsam angesehen werden, liegen die Werte hier für die historische Subskala der HCR-20 Skala zwischen 1,88 (beim Waffengebrauch) und 8,75 (Ausbruchsversuche). Der Mittelwert liegt bei 4,56 und zeigt an, dass die Skala gut brauchbar ist. Nur bei zwei von sieben verschiedenen Formen von Delikten innerhalb der Anstalt ergaben sich keine signifikanten Verbesserungen der Prognosen durch die Verwendung der betreffenden Subskala.

Es zeigen sich sehr deutliche Zusammenhänge der Skala mit Verurteilungen wegen Gewaltdelikten, Ausbruchsversuchen, Freiheitsstrafen von mindestens 30 Tagen sowie Regelverstößen im Gefängnis. Meist waren die odds für die historische Subskala deutlich höher als für die klinische Subskala (Mean Odds für die Gesamtskala 4.01; für die Historische-Skala 4.56; für die Klinische Skala 2.94).

Für die Suche nach den besten prognostischen Methoden ist zu bedenken, dass die „historischen“/biographischen Variablen, die den aktuariellen/statistischen Instrumenten unterliegen, im Prinzip

als *statisch*, d. h. als quasi unveränderlich bzw. als sehr schwer und nur langsam veränderlich angesehen werden müssen. Klinische Faktoren gelten dagegen als *dynamisch* und als durch Behandlung *veränderbar*. Da sie sich meist auf die Psychopathologie des Probanden beziehen, sind sie allerdings nur schwer valide messbar. Gegen aktuarielle Instrumente wird eingewendet, dass sie nur in Zusammenhang mit klinischer Beurteilung einsetzbar seien (Reed 1997; Price 1997) und durch *Risikomanagementvariablen* ergänzt werden müssten.

Ein weiterer vergleichender Test des VRAG und des historischen Teils des HCR-20, also des H-10, bei verschiedenen Kategorien von Rechtsbrechern, nämlich solchen mit der Diagnose Persönlichkeitsstörungen und solchen mit Schizophrenie, unter zusätzlicher Berücksichtigung der Psychopathy Checklist-Revised (PCL-R) liegt von Grann u. a. (2000) vor. Es handelt sich um eine *retrospektive* Follow Up-Studie von wegen Gewaltdelikten verurteilten Straftätern, die nach schwedischem Recht zu einer psychiatrischen Untersuchung verpflichtet waren (1988 bis 1993). Die in die Follow Up-Untersuchung einbezogenen, nicht als schizophran diagnostizierten Probanden (N = 293) sind in etwa repräsentativ für gewalttätige Straftäter mit Persönlichkeitsstörungen in schwedischen Gefängnissen. Die andere Gruppe umfasste alle schwedischen männlichen Gewalttäter mit der Diagnose Schizophrenie zwischen 1988 und 1993, die nach schwedischem Recht alle als psychisch krank galten, von denen aber aus verschiedenen Gründen nur ein Teil berücksichtigt werden konnte (N = 202). Probanden mit Gewaltdelikten (Mord, Totschlag, Körperverletzung, Raub, Vergewaltigung) innerhalb von zwei Jahren nach der Entlassung, die zu erneuter Verurteilung führten, wurden dichotom als Rückfall oder Nicht-Rückfall gezählt, andere Rückfallvarianten nicht.

Die prädiktive Validität wurde mittels der ROC-Analyse bestimmt. Die AUC-Werte wurden mit einem 95%-Vertrauensintervall und die prädiktive Validität des VRAG und des H-10 wurde für die Gesamtpopulation und die beiden Subgruppen getrennt berechnet. Bei jeder Analyse wurden der optimale Cut-Off der ROC-Kurve bestimmt und die Sensitivität, die Spezifität und der positive und der negative Prädiktionswert berechnet.

Schließt man die Psychopathie-Items aus, so ergeben sich bei Verwendung des VRAG folgende AUC-Werte: .64 bei der Gesamtkohorte, .63 bei der Kohorte mit Persönlichkeitsstörungen und .58

bei der Schizophrenie-Kohorte. Die entsprechenden Werte für das H-10 lauten .70, .69 und .66.

Tabelle 3: Prognoseleistungen des VRAG und des H (CR)-10 für verschiedene Diagnosegruppen nach Grann u. a. (2000)

	Person.Dis.+ Schizophr.		Personality Disorder		Schizophrenie	
	VRAG*	H 10**	VRAG*	H 10**	VRAG***	H 10****
AUC-Werte	.68	.71	.68	.71	.60	.66
Konfidenzintervall	.63-.73	.66-.76	.62-.73	.66-.76	.50-.69	.56-.75
Sensitivität	.50	.71	.57	.72	.68	.88
Spezifität	.77	.61	.71	.60	.53	.36
Positive Prädiktion	.39	.35	.40	.38	.20	.19
Negative Prädiktion	.84	.88	.83	.86	.91	.95

- *optimaler Cutoff bei 13 Punkten
- **optimaler Cutoff bei 12 Punkten
- ***optimaler Cutoff bei 10 Punkten
- ****optimaler Cutoff bei 8 Punkten

Bei den 404 psychisch-gestörten Patienten insgesamt ergeben die beiden Instrumente signifikant bessere Prognosen als durch Zufallszuweisung, auch wenn die Ratings *ex post* vorgenommen werden mussten, also eigentlich *keine Prognose* erfolgte. Ein Blick auf die obige, von uns konstruierte Tabelle (vgl. Tab. 3) macht anhand der AUC-Werte aber deutlich, dass beide Instrumente für alle drei Probandengruppen sehr bescheidene Prognoseleistungen erbringen, wobei die H10-Skala durchgehend etwas besser abschneidet. Bei den Probanden mit Schizophrenie unterscheidet sich die Prognoseleistung des VRAG allerdings nicht signifikant von der reinen Zufallszuweisung, hier ist das H-10-Instrument eindeutig überlegen. Die Sensitivitätswerte sind beim VRAG bei den beiden anderen Probandenklassen aber erschreckend niedrig; logischerweise sind die Spezifitätswerte deutlich besser als beim H10-Instrument.

Beide Instrumente erzielten hohe negative Prädiktionswerte (TN/(FN + TN)), d. h.: die Wahrscheinlichkeit, mit der eine negative Prognose zutreffend ist, ist hoch (Werte von .83 bis .95). Die positive Prädiktionsleistung (TP/(TP + FP)) ist aber eher bescheiden, d. h.:

Die Wahrscheinlichkeit, mit der eine positive Prognose zutreffend ist, ist relativ niedrig (Werte zwischen .19 und .40).⁷ Das bedeutet, dass maximal bei 40% der positiv prognostizierten Probanden diese Prognose zutraf.

Die diagnosespezifische Prognoseleistung macht deutlich, dass populationsspezifische Instrumente entwickelt und eingesetzt werden müssten, aber auch dann belegen die sehr niedrigen positiven Prädiktionswerte eine noch (!?) durchaus zweifelhafte Qualität der erzielten Prognosen.

4.2 Gewalt innerhalb und außerhalb der Anstalt

Bei einer statistischen Prognose über gewalttätiges Verhalten *in der Anstalt* auf der Basis vorhandener Daten (vierzehnmonatiger Prognosezeitraum) für Patienten in einer Einrichtung der Jugendpsychiatrie konnten Day u. a. (1998) mittels einer *logistischen Regression* 79% des gesamten Samples, 81% der Mädchen und 92% der Jungen, korrekt klassifizieren (signifikante Prädiktoren bei den Mädchen: Gewalt in der Familie, Minoritätenzugehörigkeit und Medikamentenbehandlung; bei den Jungen: Diagnose Verhaltensstörung, Medikamentenbehandlung und frühere Hospitalisierung). Demnach sollte ein *geschlechtsspezifisches* Instrument entwickelt werden. Bei der scheinbar guten Prädiktionsleistung ist zu bedenken, dass es um die Prognose von Gewalt *innerhalb desselben institutionellen Kontextes* ging, in dem die Probanden zum Zeitpunkt der Prognose untergebracht waren.

In einer anderen Studie (Ross u. a. 1998) fand das HCR-20 Instrument Anwendung bei psychiatrischen Patienten in akuter Kurzzeitbehandlung zur Prognose von Gewalt *im klinischen Setting* (6 Monate) und in den ersten sechs Monaten nach der Entlassung, also *in Freiheit*. Für die Zeit in der Klinik zeigte sich, dass die historischen und die klinischen Teile des HCR-20 (H-10, C-5 und HC-15) verschiedene Varianten von *Patientengewalt* mit Flächen unter der Kurve (AUC des ROC) zwischen .59 und .68 prognostizieren, wobei sich in diesem Falle die historischen Faktoren des H-10 etwas weniger gut als die klinischen Items bewährten. Bei den in die Gemeinde entlassenen Patienten prognostizierte das HCR-20 Gewaltkriminali-

⁷ Zu den Abkürzungen vgl. Fußnote 1, S. 5.

tät mit mittlerem Erfolg (AUC-Wert von .75), wobei sich die Risikomanagement-Faktoren (R-Variablen) besonders bewährten (vgl. ganz ähnliche Befunde in Schweden bei Strand u. a. 1998).

Menzies und Webster (1995) bemühten sich, die methodischen Verbesserungsvorschläge von Monahan und Steadman aufzugreifen und die sog. „Risikofaktoren“ (aktuarieller Hintergrund, klinische und rechtliche Merkmale), die „Einschätzung der zu erwartenden Risiken“ (Zuschreibung von Gefährlichkeit) und die „Schäden“ (Schwere des sich ergebenden gewalttätigen Verhaltens) bei wegen schwerer Delikte angeklagten und in die Forensik eingewiesenen Rechtsbrechern zu berücksichtigen. Die Population wies die typischen Merkmale forensischer Probanden auf (soziale Unterschicht, geringe Bildung, jung, alleinstehend, zur Hälfte frühere Psychiatrieerfahrung, dreiviertel vorbestraft, mehr als 40% mit Gefängniserfahrung, 20% wegen Gewaltdelikten). Ca. 43% wurden als mental gestört, 15% als nicht-prozesstauglich, 38% als der weiteren psychiatrischen Begutachtung, 34% als der weiteren Einsperrung und 27% als weiterer psychiatrischer Behandlung bedürftig klassifiziert.

Neben den klinischen Ratern (Psychiater, Psychologen, Sozialarbeiter, Psychiatriekrankenschwestern) wurden „unabhängige Rater“ ohne klinische Erfahrung eingesetzt, um die Prognosen von Klinikern und Nicht-Klinikern zu vergleichen.

Außer den soziodemographischen und psychosozialen Merkmalen der Probanden, die sich in anderen Studien als prognostisch relevant erwiesen hatten, kam das „Dangerous Behavior Rating Scheme“ DBRS (vgl. Megargee 1976) zur Anwendung (ursprünglich 18 Items zur *Persönlichkeit* (passive Aggression, Feindseligkeit, Ärger, Wut, Emotionalität, Schuldgefühle, Empathie, Veränderungsbereitschaft, Selbstwahrnehmung als gefährlich, Handlungskontrolle, Toleranz), zu *situationalen Gegebenheiten* (Stress in der Umwelt, Unterstützung aus dem Umfeld), zum *Lebensstil* und *Interviewverhalten des Probanden*). Des Weiteren wurden die Gefährlichkeitssteigerung durch Alkohol und/oder Drogen, manipulatives Verhalten während des Interviews, Zuverlässigkeit der Informationen im Interview etc. erfasst.

Die Interview Assessment Scale (IAS) erfasste das Verhalten der Probanden während der durch die nicht-klinischen nichtteilnehmenden Rater beobachteten Interviews (Kriterien: Grußverhalten, gepflegtes Äußeres, Blickkontaktverhalten, Affekt, körperliche Haltung, Liebenswürdigkeit, verbale Reaktionen, Kontrolle des

Interviews durch den Probanden, Schnelligkeit, Spannung, Qualität des Kontaktes, Synchronizität).

Da die Probanden in der Folgezeit durch alle möglichen Institutionen behandelt, verwahrt oder kontrolliert wurden, standen für die größte Zahl der Probanden über fast die gesamte Zeit die relevanten Informationen hinsichtlich eines Rückfalls zur Verfügung. Anhand der sorgfältigen Analyse von Akten und Registern wurde für sechs Jahre der Schaden bestimmt, den die Probanden angerichtet hatten, und zwar nach der Art des Deliktes (ob zu Wiederverurteilung führend oder nicht), dem Ort der Begehung des Deliktes (Gemeinde, Klinik oder Gefängnis) und dem zeitlichen Abstand zur Begutachtung.

Die Basisrate für Gewalttaten betrug 62%. Nach einem Jahr hatten 56 Personen (35%) eine Gesamtzahl von 158 gewalttätigen Ereignissen verursacht (0.99 Gewalthandlungen pro Kopf der Gesamtpopulation). Nach drei Jahren waren es 82 Probanden (51,6%) mit 314 gewalttätigen Ereignissen ($M = 1.97$) und nach 6 Jahren 99 Probanden (62,3%) mit 456 derartigen Ereignissen ($M = 2.87$). 60 Probanden hatten kein einziges Gewaltdelikt mehr, 32 Probanden hatten ein Delikt, 18 Probanden hatten zwei, 25 hatten drei, 15 hatten sechs bis zehn derartige Straftaten und zehn Probanden hatten mehr als zehn neue Gewaltdelikte begangen. Die meisten Gewaltereignisse ereigneten sich in der Gemeinde, sehr viel weniger im Hospital und noch weniger im Gefängnis.

Sehen wir von der Bedeutsamkeit der diversen Probandenmerkmale für die Prognoseerstellung der Kliniker und Nicht-Kliniker hier ab und betrachten nur, welche Merkmale für die Erklärung der tatsächlich eingetretenen Gewalthandlungen bedeutsam sind, so erwiesen sich in einer multiplen Regression für drei der vier Varianten der tatsächlichen Gewalt (*Gewalt insgesamt*, *Gewalt in der Gemeinde* und *Gewalt im Gefängnis*) die *sozio-demographischen* und *rechtlichen* Merkmale insgesamt als statistisch *nicht* signifikant (gerade noch signifikant für die Gewalt *innerhalb der Forensik*). Gewalt *über alle Kontexte* hinweg war noch am besten prognostizierbar, doch besaß auch dieses Modell *praktisch keine Relevanz*, da es nur 20% der Varianz erklärte.

Während sich auf *bivariater* Ebene keines der *DBRS-Items* mit dem tatsächlichen Rückfallverhalten in Zusammenhang stehend zeigte, ergibt sich bei *multipler Regression* von Rückfall (Gewalt in der Gemeinde; Gesamtgewalt; Gewalt im Hospital) auf die DBRS-

Items selbst beim besten Modell (Gewalt in psychiatrischer Klinik) nur eine Varianzaufklärung von nicht mehr als 20%) (IAS-Skala entsprechend).

Erschreckend sind die Befunde, dass die Variablen, die für das *tatsächliche Auftreten* späterer Gewalt kausal sind, *nicht* diejenigen sind, die für die *Risikoeinschätzung* durch die Prognostiker relevant sind (zwischen der Risikoprognoze und der tatsächlichen Rückfallgewalt bestehen praktisch keine relevanten Beziehungen) und dass die Prognosen der Kliniker insgesamt nicht besser waren als die der Nicht-Kliniker; eher im Gegenteil.

Insgesamt war die Prognose offensichtlich ein einziger Reinfluss, obwohl die Rückfallrate (Basisrate) so hoch war, dass es eigentlich zu besseren Prognosen hätte kommen können. Es zeigte sich aber auch nur eine sehr schwache Beziehung zwischen vermeintlichen aktuariellen und klinischen Risikofaktoren auf der einen und tatsächlichen Schäden durch Rückfall auf der anderen Seite, wodurch Autoren, die die großen Probleme der Prognose von Gewalt hervorgehoben hatten, bestätigt werden.

5. Schlussfolgerungen

Die hier für die Prognose von Gewalt durch psychisch Kranke berichteten Befunde, die mit denen für Gewaltkriminalität insgesamt übereinstimmen (vgl. Albrecht 2003a, 2003b), müssen sehr nachdenklich stimmen. Das Interesse der Öffentlichkeit, vor Gewalttaten geschützt zu werden, ist verständlich und berechtigt und damit auch der Versuch, diesem Ziel durch Prognosen ein wenig näher zu kommen. Der Gesetzgeber hat in Deutschland aber ganz offensichtlich unrealistische Erwartungen an die Leistungsfähigkeit der Gefährlichkeitsprognose. So hat er die Bestimmungen zum Einsatz von Prognosen vor fünf Jahren in einigen Punkten drastisch verschärft. Die höchsten Anforderungen an die Sicherheit der Prognose werden bei der Entlassung aus dem Maßregelvollzug gestellt. Die Aussetzungsklausel für die Maßregeln in § 67 d Abs. 2 StGB lautet seit dem 31.01.1998: „Ist keine Höchstfrist vorgesehen oder ist die Frist noch nicht abgelaufen, so setzt das Gericht die weitere Vollstreckung der Unterbringung zur Bewährung aus, wenn zu erwarten ist, dass der Untergebrachte außerhalb des Maßregelvollzugs keine rechtswidrigen Taten mehr begehen wird“,

während die alte Regelung „nur“ vorsah, dass dies geschehen sollte „sobald verantwortet werden kann zu erproben, ob der Untergebrachte außerhalb des Maßregelvollzugs keine rechtswidrigen Taten mehr begehen wird“. Eine entsprechende Verschärfung erfahren auch die Aussetzungsklauseln für Strafreste (§ 57 Abs. 1 Nr. 2 StGB). Während zuvor eine Aussetzung unter bestimmten Bedingungen möglich war, um zu erproben, ob der Verurteilte keine neuen Straftaten mehr begeht, ist nun die Voraussetzung, dass der Entscheider prognostiziert, dass der Untergebrachte/Verurteilte *keine* rechtswidrigen Straftaten mehr begehen wird, die Wahrscheinlichkeit weiterer Straftaten also praktisch gleich null sein muss (ähnlich bei der Entlassung aus der Haftstrafe und aus dem Maßregelvollzug).

Angesichts der oben dargelegten, ganz erheblichen Fehleranfälligkeit von Prognosen müsste der Prognostiker eigentlich *alle* entsprechenden Probanden für gefährlich erklären, wenn er nicht in einem Sinne irren will, den ihm die Öffentlichkeit oder die Justiz vorwerfen könnte. Wenn der Gesetzgeber dem Prognostiker bei seiner statistischen Prognose nicht bestimmte, nicht zu gering zu veranschlagende Irrtumswahrscheinlichkeiten einräumt, kann dieser nur durch eine äußerst restriktive/repressive Prognosepraxis reagieren. Dann aber stellt sich die Frage, wer uns vor der Gefährlichkeit der Gefährlichkeitsprognostiker schützt, denn die Prognostiker stehen selbst durch diese Gesetzgebung und die Hysterie in der Öffentlichkeit unter schwerstem Druck (vgl. Nedopil 1999), teilweise mit fatalen Folgen (vgl. Volckart 1998; Albrecht 2003b).

Literatur

- Albrecht, Günter (1990): Möglichkeiten und Grenzen der Prognose „krimineller Karrieren“, in: Schriftenreihe der DVJJ, Bd. 18, Mehrfach Auffällige – Mehrfach Betroffene. Erlebnisweisen und Reaktionsformen. Bonn-Bad Godesberg: Forum-Verlag, S. 99–116.
- Albrecht, Günter (1999): Methodische Probleme der Erforschung sozialer Probleme, in: Günter Albrecht/Axel Groenemeyer/Friedrich Stallberg (Hrsg.): Handbuch soziale Probleme. Opladen/Wiesbaden: Westdeutscher Verlag, S. 768–882.
- Albrecht, Günter (2003a): Sinn und Unsinn der Prognose von Gewalt. Erscheint in: Wilhelm Heitmeyer/Hans Georg Soeffner (Hrsg.): Paradigmen der Gewaltforschung. Frankfurt am Main: Suhrkamp.

- Albrecht, Günter (2003b): Die Prognose von Gewalt. Theoretische, methodische und empirische Analysen. Preprint der Fakultät für Soziologie der Universität Bielefeld. Bielefeld.
- Amelang, Manfred/Zielinski, Werner (1997): Psychologische Diagnostik und Intervention, 2. korrig., aktual. und erw. Aufl. Berlin-Heidelberg-New York: Springer.
- Andrews, Dan A. (1989): Recidivism Is Predictable and Can Be Influenced: Using Risk Assessment To Reduce Recidivism. Forum On Corrections Research, 1, pp. 11–18.
- Black, Tony/Spinks, Penny (1985): Predicting Outcomes of Mentally Disordered and Dangerous Offenders, in: David P. Farrington/Roger Tarling (Eds.): Prediction in Criminology. Albany: State University of New York Press, pp. 174–191.
- Clear, Todd (1988): Statistical Prediction in Corrections. Research in Corrections, 1 (1), pp. 1–40.
- Cocozza, Joseph J. (1975): From Deviant to „Normal“: Factors Associated With the Official Delabeling of Criminally Insane Patients. Dissertation Case Western Reserve University.
- Cocozza, Joseph J./Steadman, Henri J. (1976): The Failure of Psychiatric Predictions of Dangerousness: Clear and Convincing Evidence. Rutgers Law Review, 29, pp. 1084–1101.
- Day, H.D./Franklin, J.M./Marshall, D.D. (1998): Predictors of Aggression in Hospitalized Adolescents. The Journal of Psychology, 132 (4), pp. 427–434.
- Douglas, Kevin S./Webster, Christopher D. (1999): The HCR-20 Violence Risk Assessment Scheme. Concurrent Validity in a Sample of Incarcerated Offenders. Criminal Justice and Behavior, 26 (1), pp. 3–19.
- Farrington, David P./Tarling, Roger (1985a): Criminological Prediction: An Introduction, in: David P. Farrington/Roger Tarling (Eds.): Prediction in Criminology. Albany, N.Y.: State Univ. of New York Press, pp. 2–33.
- Farrington, David P./Tarling, Roger (1985b): Criminological Prediction: The Way Forward, in: David P. Farrington/Roger Tarling (Eds.): Prediction in Criminology. Albany, N.Y.: State University of New York Press, pp. 258–269.
- Faust, David/Nurcombe, Barry (1989): Improving the Accuracy of Clinical Judgment. Psychiatry, 52, pp. 197–208.
- Gabor, Thomas (1986): The prediction of criminal behavior. Statistical approaches, Toronto: Univ. of Toronto Press.
- Gottfredson, Stephen D. (1987): Statistical and Actuarial Considerations, in: Ferdinand Dutilleul/Cleon Foust (Eds.): The Prediction of Violence. Springfield, Ill.: Charles C. Thomas, pp. 71–81.
- Grann, Martin/Belfrage, Henrik/Tengström, Anders (2000): Actuarial Assessment of Risk For Violence. Predictive Validity of the VRAG and the Historical Part of the HCR-20. Criminal Justice and Behavior, 27 (1), pp. 97–114.
- Hare, Robert D. (1991): Manual for the Hare Psychopathy Checklist Revisited, Toronto: Multi-Health Systems.
- Harris, Grant T./Rice, Marnie E. (1997): Risk Appraisal and Management of Violent Behavior. Psychiatric Services, 48, pp. 1168–1176.
- Harris, Grant T./Rice, Marnie E./Cormier, Catherine A. (1991): Psychopathy and Violent Recidivism. Law and Human Behavior, 15, pp. 625–637.

- Harris, Grant T./Rice, Marnie E./Quinsey, Vernon L. (1993): Violent Recidivism of Mentally Disordered Offenders. The Development of a Statistical Prediction Instrument. *Criminal Justice and Behavior*, 20 (4), pp. 315–335.
- Hodgins, Sheilag (1990): Prevalence of mental disorders among penitentiary inmates in Quebec. *Canada's Mental Health*, 38, pp. 1–4.
- Holland, Terril R./Holt, Norman/Levi, Mari/Beckett, Gerald E. (1983): Comparison and Combination of Clinical and Statistical Predictions of Recidivism Among Adult Offenders. *Journal of Applied Psychology*, 68, pp. 203–211.
- Koehler, Karl/Saß, Henning (Hrsg.) (1984): Diagnostisches und statistisches Manual psychischer Störungen DSM III. Deutsche Bearbeitung. Weinheim-Basel: Beltz.
- Kühl, Jörn/Schumann, Karl F. (1989): Prognosen im Strafrecht – Probleme der Methodologie und Legitimation. *Recht und Psychiatrie*, 7, S. 126–148.
- Lab, Steven P. (1988): *Crime Prevention: Approaches, Practices and Evaluations*. Cincinnati, Ohio: Anderson.
- Lefrenz, Heinz (1972): Neuere Ergebnisse der Kriminalprognostik bei Jugendlichen, in: Heinrich Koch/Hermann Stutte (Hrsg.): *Jugenddissozialität*. Villingen: Neckar-Verlag, S. 57–63.
- Leygraf, Norbert/Nowara, Sabine (1992): Prognosegutachten, Klinisch-psychiatrische und psychologische Beurteilungsmöglichkeiten der Kriminalprognose. *Forensia-Jahrbuch*, 3, S. 43–53.
- Loeber, Rolf/Stouthamer-Loeber, Magda (1987): Prediction, in: Herbert C. Quay (Ed.): *Handbook of Juvenile Delinquency*. New York: Wiley, pp. 325–382.
- Lösel, Friedrich (1982): Prognose und Prävention von Delinquenzproblemen, in: Jochen Brandstädter/Alexander von Eye (Hrsg.): *Psychologische Prävention*. Bern: Huber, S. 97–239.
- Megargee, Edwin I. (1976): The Prediction of Dangerous Behavior. *Criminal Justice and Behavior*, 3, pp. 3–22.
- Menzies, Robert/Webster, Christopher D. (1995): Construction and Validation of Risk Assessments in a Six-Year Follow-Up of Forensic Patients – A Tridimensional Analysis. *Journal of Consulting and Clinical Psychology*, 63 (5), pp. 766–778.
- Monahan, John (1975): The Prediction of Violence, in: Duncan Chappell/John Monahan (Eds.): *Violence and Criminal Justice*. Lexington/Toronto/London: Lexington, pp. 15–31.
- Monahan, John (1981): *Predicting Violent Behavior. An Assessment of Clinical Techniques*. Beverly Hills: Sage.
- Monahan, John (1984): The Prediction of Violent Behavior: Toward a Second Generation of Theory and Policy. *American Journal of Psychiatry*, 141, pp. 10–15.
- Monahan, John (1992): Mental Disorder and Violent Behavior. *American Psychologist*, 47, pp. 511–521.
- Monahan, John (1996): Violence Prediction. The Past Twenty and the Next Twenty Years. *Criminal Justice and Behavior*, 23 (1), pp. 107–120.
- Monahan, John/Steadman, Henry J. (1994): Toward a Rejuvenation of Risk Assessment Research, in: John Monahan/Henry J. Steadman (Eds.): *Violence and mental disorder*. Chicago/London: Univ. of Chicago Press, pp. 1–17.
- Mossman, Douglas (1994): Assessing Predictions of Violence: Being Accurate about Accuracy. *Journal of Consulting and Clinical Psychology*, 62, pp. 783–792.
- Mummendey, Hans Dieter (1981): Methoden und Probleme der Kontrolle sozialer Erwünschtheit. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 2, S. 199–218.
- Nedopil, Norbert (1999): Begutachtung zwischen öffentlichem Druck und wissenschaftlicher Erkenntnis. *Recht und Psychiatrie*, 17, S. 120–126.
- Nedopil, Norbert (unter Mitarbeit von Volker Dittmann, Franz Joseph Freisleder und Reinhard Haller) (2000): *Forensische Psychiatrie. Klinik, Begutachtung und Behandlung zwischen Psychiatrie und Recht*, 2., aktualisierte und erweiterte Aufl. Stuttgart/New York: Thieme.
- Nuffield, John (1982): *Parole Decision-Making in Canada: Research Towards Decision Guidelines*. Ottawa: Ministry of Supply and Services.
- Phillips, Derek L. (1973): *Abandoning Method. Sociological Studies on Methodology*. San Francisco/Washington/London: Jossey-Bass.
- Price, R. (1997): On the Risks of Risk Predictions. *The Journal of Forensic Psychiatry*, 8, pp. 1–4.
- Quinsey, Vernon L. (1980): The Base-rate Problem and the Prediction of Dangerousness: A Reappraisal. *Journal of Psychiatry and Law*, 8, pp. 329–340.
- Quinsey, Vernon L. (1995): The Prediction and Explanation of Criminal Violence. *International Journal of Law and Psychiatry*, 18, pp. 117–127.
- Quinsey, Vernon L./Harris, Grant T./Rice, Marnie E./Cormier, Catherine A. (1998): *Violent Offenders. Appraising and Managing Risk*. Washington, D. C.: American Psychological Association.
- Reed, John (1997): Risk Assessment and Clinical Risk Management: The Lessons From Recent Inquiries. *British Journal of Psychiatry*, 170, Supplement 32, pp. 4–7.
- Rice, Marnie E. (1997): Violent Offender Research and Implications for the Criminal Justice System. *American Psychologist*, 52, pp. 414–423.
- Rice, Marnie E./Harris, Grant T. (1995): Violent Recidivism: Assessing Predictive Validity. *Journal of Consulting and Clinical Psychology*, 63, pp. 737–748.
- Rice, Marnie E./Harris, Grant T. (1997): Cross-Validation and Extension of the Violence Risk Appraisal Guide for Child Molesters and Rapists. *Law and Human Behavior*, 21, pp. 231–241.
- Ross, D. J./Hart, S. D./Webster, C. D. (1998): *Facts and Fates: Testing the HCR-20 Against Aggressive Behavior in Hospital and Community*, Vancouver, Can.
- Schultz, Peter (1975): *Zum Problem der Prognose in der Bewährungshilfe*. Dissertation Universität Köln.
- Spiß, Gerhard (1993): *Kriminalprognose*, in: Günther Kaiser/Hans-Jürgen Kerner/Fritz Sack/Hartmut Schellhoss (Hrsg.): *Kleines Kriminologisches Wörterbuch*, 3. Aufl. Heidelberg: C. F. Müller, S. 286–294.
- Steadman, Henri J. (1987): How Well Can We Predict Violence For Adults? A Review of the Literature and Some Commentary, in: Ferdinand N. Dutilleul/Cleon H. Foust (Eds.): *The Prediction of Criminal Violence*. Springfield, Ill.: Charles C. Thomas, pp. 5–19.
- Steadman, Henri J./Silver, Eric/Monahan, John/Appelbaum, Paul S./Robbins, Pamela Clark/Mulvey, Edward P./Grisso, Thomas/Roth, Loren H./Banks, Steven (2000): *Enhancing the Clinical Usefulness of Actuarial Violence Risk Assessment Tools: A Classification Tree Approach*. *Law and Human Behavior*, 24, pp. 83–100.

- Steadman, Henry J./Cocozza, Joseph (1974): *Careers of the Criminally Insane: Excessive Social Control of Deviance*. Lexington, Mass: Lexington.
- Steadman, Henry J./Cocozza, Joseph (1978): *Psychiatry, Dangerousness and the Repetitively Violent Offender*. *The Journal of Criminal Law and Criminology*, 69 (2), pp. 226–231.
- Steadman, Henry J./Mulvey, Edward P./Monahan, John/Robbins, Pamela Clark/Appelbaum, Paul S./Grisso, Thomas/Roth, Loren H./Silver, Eric (1998): *Violence by People Discharged From Acute Psychiatric Inpatient Facilities and by Others in the Same Neighborhoods*. *Archiv Gen. Psychiatry*, 55, pp. 393–401.
- Strand, Susan/Belfrage, Henrik/Fransson, Goran/Levander, Sten (1998): *Clinical and Risk Management Factors in Risk Prediction of Mentally Disordered Offenders – More Important Than Actuarial Data? A Retrospective Study of 40 Mentally Disordered Offenders Assessed With the HCR-20 Violence Risk Assessment Scheme*. *Legal and Criminal Psychology*, 4, pp. 67–76.
- Tardiff, Kenneth (1989): *A Model for the Short-Term Prediction of Violence Potential*, in: David A. Brizer/Martha Crowner (Eds.): *Current Approaches to the Prediction of Violence*. Washington, D. C.: American Psychiatric Press, pp. 3–12.
- Tarling, Roger/Perry, John A. (1985): *Statistical Methods in Criminological Prediction*, in: David P. Farrington/Roger Tarling (Eds.): *Prediction in Criminology*. Albany, N. Y.: New York State University Press, pp. 210–231.
- Volckart, Bernd (1998): *Die Aussetzungsprognosen nach neuem Recht*. *Recht und Psychiatrie*, 16, S. 3–11.
- Volckart, Bernd (1999): *Zur Bedeutung des hermeneutischen Verstehens in der Kriminalprognose*. *Recht und Psychiatrie*, 17, S. 58–64.
- Weber, Hartmut (1988): *Gefährlichkeitsprognosen bei „Lebenslänglichen“ – Gefahr für wen?*, in: Hartmut Weber/Sebastian Scheerer (Hrsg.): *Leben ohne lebenslänglich*. Bielefeld: AJZ, S. 85–123.
- Webster, C. D./Harris, Grant T./Rice, Marnie E./Cormier, Catherine A./Quinsey, Venon L. (1994): *The Violence Prediction Scheme. Assessing Dangerousness in High Risk Men*. Toronto: Centre of Criminology.

Der Autor:

Prof. Dr. Günter Albrecht
 Universität Bielefeld
 Fakultät für Soziologie und
 Institut für interdisziplinäre Konflikt- und Gewaltforschung
 Universitätsstr. 25
 33615 Bielefeld

E-Mail: günter.albrecht@uni-bielefeld.de